ISSN: 2706-6495

الطلبة ذوو الاستجابات الشاذة في اختبار الرياضيات TIMSS2019 بسلطنة عمان (مقارنة أداء مؤشر مربع البواقي البينية حسب المحتوى والعمليات المعرفية ونوع الأسئلة وصعوبتها)

Students with Aberrant Responses on TIMSS2019 Mathematics Test in Sultanate of Oman: Comparison of Between Squared Residual Index by Content, Cognitive Process, Item Type, and Difficulty

إعداد: الباحثة/ نوال علي اليحيائي

ماجستير قياس وتقويم، جامعة السلطان قابوس، سلطنة عُمان

Email: nawal.alyahyai@gmail.com

أ.د/ راشد سيف المحرزي

أستاذ دكتور في القياس والتقويم، قسم علم النفس، كلية التربية، جامعة السلطان قابوس، سلطنة عُمان

الدكتور/ إيهاب محمد عمارة

أستاذ مساعد، قسم علم النفس، كلية التربية، جامعة السلطان قابوس، سلطنة عُمان

الملخص:

هدفت هذه الدراسة إلى الكشف عن الطلبة ذوي الاستجابات غير المطابقة عبر أقسام اختبار الرياضيات في TIMSS2019 وفقا لأربعة مجالات: المحتوى، العمليات المعرفية، نوع الأسئلة، صعوبة الأسئلة. وطبق الاختبار على طلبة الصف الثامن بسلطنة عمان في عام 2019، والبالغ عددهم 6745 طالبا وطالبة، وذلك باستخدام مؤشر مربع البواقي البينية Between Squared Residual Index، كما هدفت هذه الدراسة إلى مقارنة نتائج مؤشر مربع البواقي البينية على مستوى أقسام اختبار الرياضيات بنتائج استخدام مؤشر مربع البواقي على مستوى المفردة Squared Residual Index. وأشارت نتائج الدراسة إلى أن نسبة الأفراد ذوي الاستجابات غير المطابقة تراوحت بين 2.16% و3.22% حسب مؤشر مربع البواقي البينية غير الموزون، وبين2,13 % و3.88% حسب مؤشر مربع البواقي البينية الموزون، بينما بلغت نسبتهم 0.30% فقط وفقا لمؤشر مربع البواقي غير الموزون، و0.95% وفقا لمؤشر مربع البواقي الموزون، وأظهرت النتائج أن أغلب قدرات الأفراد ذوي الاستجابات غير المطابقة في اختبار الرياضيات في TIMSS2019 تراوحت بين المنخفضة والمتوسطة. وأوصت الدراسة بإجراء دراسات للوقوف على أسباب وجود الاستجابات غير المطابقة في اختبارات الرياضيات في TIMSS لدى طلبة السلطنة، كما أوصت بتوجيه المعلمين والمعلمات والقائمين على تطبيق اختبارات TIMSS بضرورة تدريب الطلبة على المهارات الرياضية الأساسية التي يحتاجونها لأداء هذه الاختبارات للتقليل من لجوئهم إلى استخدام سلوكيات تؤدي إلى استجابات غير مطابقة.

الكلمات المفتاحية: الاستجابات غير المطابقة، مؤشر مربع البواقي البينية، مؤشر مربع البواقي، اختبار الرياضيات في TIMSS2019

ISSN: 2706-6495

Students with Aberrant Responses on TIMSS2019 Mathematics Test in Sultanate of Oman (Comparison of Between Squared Residual Index by Content, Cognitive Process, Item Type, and Difficulty)

ABSTRACT:

This study aimed to detect the person misfit for eight grade students in the Sultanate of Oman using the Between Squared Residual Index over sections of TIMSS2019 Mathematics test (content, cognitive levels, item format and item difficulty). This study also aimd to compare the results of the Between Squared Residual Index over Mathematics test sections with the results of the Squared Residual Index over item level. The results of the study indicated that the percentage of students with misfit responses ranged between (2.16% & 3.22%) and between (2.13% & 3.88%) according to the (unweighted and weighted) Between Residual Square Indices respectively, while their percentage amounted to only 0.30% & 0.95% according to the (unweighed & weighed) Square Indices respectively. The results showed that most of students ' abilities with misfit responses in the Math test in TIMSS2019 ranged between low and medium. The study recommended conducting studies to determine the reasons for the presence of misfit responses in Mathematics tests in TIMSS among students of the Sultanate. It also recommended directing teachers and those in charge of applying TIMSS tests to train students in the basic mathematical skills they need to reduce behaviors that lead to misfit responses.

Keywords: Person misfit, between squared residual index, squared residual index, TIMSS Mathematics test

ISSN: 2706-6495

1. المقدمة

نظرا لأهمية التعليم باعتباره أساسا للنمو والتنافسية بين الدول تحرص الدول المختلفة على تطوير أنظمة التعليم لديها، وذلك من خلال مقارنة أدائها بأداء أنظمة تعليمية مختلفة، ويكون ذلك من خلال التقييمات الدولية التي توفر ثروة من البيانات التي تسهم في تحسين السياسات التعليمية. ومن هذا المنطلق حرصت وزارة التربية والتعليم في سلطنة عمان على المشاركة في الاختبارات الدولية ومنها اختبارات الاتجاهات العالمية في التحصيل الدراسي للرياضيات والعلوم Trends of the International Mathematics and Science Studies (TIMSS) والتي تعتبر أحد المؤشرات في قياس جودة التعليم في مادتي العلوم والرياضيات. وتعد اختبارات TIMSS في العلوم والرياضيات جزءا من مجموعة من الأدوات التي تهدف للحصول على بيانات شاملة عن المفاهيم والمواقف التي تعلمها الطلبة في مادتي العلوم والرياضيات في الصفين الرابع والثامن. وتهدف هذه الاختبارات إلى تقييم التحصيل، وتوفير المعلومات لتحسين تعليم وتعلم الرياضيات والعلوم في الدول المشاركة. وتتم هذه الدراسة تحت إشراف الجمعية الدولية لتقييم التحصيل التربوي International Association for the Evaluation of Educational Achievement (IEA) ومقرها في مدينة أمستردام الهولندية، وتعمل هذه المنظمة على رصد نتائج هذا الاختبار الذي تشارك فيه أكثر من 60 دولة، ويعقد دوريا كل 4 سنوات (زكي، 2011).

وتعتمد القرارات السليمة في الأوساط التعليمية إلى حد كبير على القياس الدقيق لخصائص الطالب، حيث يساعد ذلك صانعي السياسات على اتخاذ قرار بشأن القضايا التعليمية الهامة. في المقابل، يمكن أن يؤدي القياس غير الدقيق لأداء الاختبار إلى عواقب سلبية، حيث يمكن أن تؤدي درجات الاختبار المرتفعة بشكل زائف إلى تقويض تقييم تقدم تعلم الطلاب وجهود تخطيط المناهج الدراسية (Karabatoss, 2003).

وقد اهتم علماء القياس منذ أن وجدت حركة القياس النفسي بالطرق التي من شأنها تحقيق أعلى درجة من الموضوعية في أدوات القياس، وقد أسفرت جهود العلماء عن ظهور بعض الاتجاهات الحديثة في مجال القياس والتقويم، ومن بين هذه الاتجاهات نظرية الاستجابة للمفردة Item Response Theory (IRT) التي أضافت الكثير من الطرق لتحقيق درجات عالية من الدقة فيما يتعلق بالحكم على نتائج أدوات القياس.

ومن الطرق التي أوجدتها نظرية الاستجابة للمفردة مؤشر مطابقة الفرد Person Fit Statistic (PFS) والذي يساعد في تحديد أنماط الاستجابة غير المطابقة أو الشاذة. وتكون الاستجابات للأفراد غير مطابقة حين يختلف نمط إجابتهم بشكل كبير عن نمط الإجابة المتوقع، وبالتالي تكون لاستجاباتهم قيم متطرفة وفقا لمؤشر مطابقة الفرد المستخدم، ومن ثم يحدد نمط إجابتهم على أنه غير متسق أو غير مطابق لنموذج IRT المستخدم (Ferrando & Chico, 2001) .

وتحدث الاستجابات غير المطابقة نتيجة عدة عوامل تؤثر على أداء الفرد، ومنها كما ذكر Wright(1977) التكاسل نتيجة الملل مما يؤدي إلى أن يجيب الفرد إجابات غير دقيقة خصوصا في الأسئلة الأخيرة للاختبار، والارتباك الذي يحدث بسبب شكل الاختبار والذي يؤدي إلى إجابات غير دقيقة في الأسئلة الأولى من الاختبار، والتباطؤ في حل الأسئلة مما يفوت حل الأسئلة الأخيرة، والتخمين، والغش. وتحدث الاستجابات غير المطابقة أو الشاذة أيضا بسبب الإبداع الاستثنائي لدى بعض الطلبة في الإجابة على الأسئلة. كما يظهر بعض الأفراد أيضا استجابات غير مطابقة كأن يكونوا ذوي قدرات عالية، أو تلقوا تعليما بغير لغتهم الأم (Bracey et al., 1992).

وقد اهتمت العديد من الدراسات بالبحث في أسباب ظهور الاستجابات غير المطابقة للأفراد في الاختبارات المختلفة، ومنها دراسة بني عطا (2019) التي أشارت إلى أن التخمين والغش والتكاسل والإبداع الاستثنائي كانت من أسباب ظهور أنماط الاستجابة غير المطابقة في اختبار أوتيس- لينون المقنن للبيئة الأردنية، ودراسة برنبوم (Birenbaum, 1986) التي أشارت إلى تأثير القلق على أداء الأفراد في اختبار القدرات، ودراسة تشن (Chen, 2004) التي أشارت إلى تأثير القلق وضغط الوقت على أداء ذوي القدرات المتوسطة والمرتفعة في اختبار لغة إنجليزية طبق لدى الطلبة الصينيين. كما أظهرت دراسة براون وفيلاريل (Brown & Villareal, 2007) أن الغش كان له دور في ظهور استجابات غير مطابقة لدى الطلبة منخفضي التحصيل في اختبار تكيفي في مادة الرياضيات.

مؤشرات مطابقة الفرد

وتساعد مؤشرات مطابقة الفرد في التعرف على الأنماط غير المطابقة كأن يكون الطالب ذي قدرة متدنية ويجيب على أسئلة ذات مستوى عال أو العكس. ويرى ميجر (Meijer, 2003) أن مطابقة الفرد مهمة لتحديد ما إذا كان المفحوص قد أجاب وفقا للسمة الأساسية التي يتم قياسها أو إذا كانت هناك عوامل أخرى قد أثرت على سلوك الاستجابة (Steinkamp, 2017).

وهناك نوعان من مؤشرات مطابقة الفرد: 1) على مستوى المفردة الواحدة و2) على مستوى أقسام الاختبار. وهناك عدد من المؤشرات على مستوى المفردة تم استخدامها من قبل المختصين في القياس النفسي والتربوي وقد بلغ عددها كما ورد في دراسة كاراباتوس(2003) Karabatoss 36 مؤشرا. ومن أمثلة المؤشرات على مستوى المفردة مؤشر رايت (1977) Wright المبني على البواقي ومؤشر درازجو L_z Drasgow et al. (1985) الذي يعتمد على الأرجحية العظمى ومؤشر مربع البواقي للمحرزي (2003) Almehrizi المبني على البواقي أيضا.

وهناك مؤشران من ضمن مؤشرات مطابقة الفرد على مستوى أقسام الاختبار: مؤشر Wright (1977)، ومؤشر مربع البواقي البينية Between Residual Square Index والذي قام المحرزي (Almehrizi, 2019) بتطويره وهو يتمتع بخصائص إحصائية جيدة كما جاء في نتائج دراسته والتي قام فيها بدراسة الخصائص الإحصائية لمؤشري مربع البواقي البينية الموزون وغير الموزون، ومؤشري (Wright, 1977)الموزون وغير الموزون.

ويمكن التفريق بين مؤشرات مطابقة الفرد على مستوى المفردة وبين تلك على مستوى أقسام الاختبار في أن الأولى يتم فيها حساب قيمة المؤشر من درجات كل مفردة من مفردات الاختبار التي يحصل عليها الطالب، بينما يتم حساب قيمة المؤشر في المؤشرات على مستوى أقسام الاختبار من درجات الطالب على كل قسم من أقسام الاختبار. وتبحث المؤشرات على مستوى أقسام الاختبار ظهور الاستجابات الشاذة أو غير المطابقة في الدرجات الفرعية لأقسام الاختبار من خلال مقارنتها بالدرجات الفرعية المتوقعة لأقسام الاختبار حسب نموذج نظرية الاستجابة للمفردة المستخدم.

مؤشر مربع البواقي البينية Between Squared Residual Index

يعتبر مؤشر مربع البواقي البينية من المؤشرات الحديثة على مستوى أقسام الاختبار، وسيتم التطرق لمؤشر مربع البواقي على مستوى المفردة باعتباره الأساس الذي انطلق منه مؤشر مربع البواقي البينية.

يعتبر مؤشر مربع البواقي على مستوى المفردة Squared Residual Index من المؤشرات التي توظف منهج البواقي بين استجابة الفرد الملاحظة واحتمال الإجابة الصحيحة على المفردة وهو بنسختين: غير موزونة USR_a وموزونة WSR_a ويتم حساب غير الموزونة كما يلي:

ISSN: 2706-6495

$$USR_a = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{(y_{ai}-p_{ai})^2-p_{ai}q_{ai}}{\sqrt{p_{ai}q_{ai}(p_{ai}-q_{ai})^2}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{(p_{ai}-y_{ai})(p_{ai}-q_{ai})}{\sqrt{p_{ai}q_{ai}(p_{ai}-q_{ai})^2}}$$

حيث y_{ai} تمثل استجابة الفرد a على المفردة i من مفردات الاختبار وعددها n، p_{ai} تمثل احتمال إجابة الفرد a على المفردة i، q_{ai} تمثل $(1-p_{ai})$.

أما النسبة الموزونة فيتم حسابها كالتالي:

$$WSR_a = \frac{\sum_{i=1}^{n}(y_{ai}-p_{ai})^2-\sum_{i=1}^{n}p_{ai}q_{ai}}{\sqrt{\sum_{i=1}^{n}p_{ai}q_{ai}(p_{ai}-q_{ai})^2}} = \frac{\sum_{i=1}^{n}(p_{ai}-y_{ai})(p_{ai}-q_{ai})}{\sqrt{\sum_{i=1}^{n}p_{ai}q_{ai}(p_{ai}-q_{ai})^2}}.$$

ويذكر (Almehrizi, 2010) أنه إذا جاءت البيانات ملائمة للنموذج الرياضي لنظرية الاستجابة للمفردة فإن كلا من USR_a وWSR_a يتبع التوزيع الطبيعي الاعتدالي بمتوسط حسابي صفر وانحراف معياري واحد، وبالتالي فإن قيم مؤشرات USR_a و WSR_a ستكون كبيرة وموجبة لتظهر قيما غير مطابقة في أنماط الاستجابة، وهي القيم التي تزيد عن القيمة 1.645 والتي تناظر مستوى دلالة 0.05 (عودة، 2019).

ومن الدراسات التي تناولت مؤشر مربع البواقي على مستوى المفردة دراسة المحرزي (Almehrizi, 2010) التي قارن فيها الخصائص الإحصائية لمؤشري مربع البواقي (الموزون وغير الموزون) بالخصائص الإحصائية لمؤشري رايت (Wright, 1977) الموزون وغير الموزون باستخدام بيانات مولدة تكونت من اثنتي عشر مجموعة من البيانات نتجت عن اختبارين مختلفين في الطول، وثلاثة مستويات مختلفة في الصعوبة، ومستويين مختلفين في التمييز. وأظهرت النتائج تمتع مؤشري مربع البواقي بخصائص إحصائية أفضل تحت الظروف الاختبارية المختلفة. كما هدفت دراسة الشقصي (2019) إلى فحص فعالية ثلاثة مؤشرات لمطابقة الفرد (رايت الموزون ودرازجو و مربع البواقي الموزون) في نماذج استجابة الفقرة عند اختلاف قوة الارتباط الموضعي بين الفقرات، وأشارت النتائج بشكل مباشر إلى أن نسبة الأفراد غير المطابقين تزداد بزيادة الارتباط الموضعي بين الفقرات للمؤشرات الثلاثة (رايت الموزون، ودرازجو، ومربع البواقي الموزون)، وأشارت الإحصاءات الوصفية والالتواء والتفلطح لمؤشرات مطابقة الفرد الى أن قيم مؤشر مربع البواقي الموزون تتبع التوزيع الطبيعي الاعتدالي في حالة الارتباط الموضعي 0.0، وتتفق هذه النتيجة مع ما ورد في دراسة (Almehrizi, 2010) والتي أشار فيها إلى أنه إذا طابقت البيانات النموذج المستخدم فإن نتائج المؤشر تتبع التوزيع الطبيعي بمتوسط حسابي صفر وانحراف معياري واحد، أما في حالات الارتباط الموضعي 0.3 و 0.6 و 0.9 فقد ابتعدت النتائج عن التوزيع الطبيعي وحافظت على شكل التوزيع.

وسعت دراسة عودة (2019) إلى مقارنة فعالية خمسة مؤشرات لمطابقة الفرد على مستوى المفردة: مؤشر مربع البواقي الموزون WSR وغير الموزون USRومؤشر درازجو L_z ورايت الموزون WT وغير الموزون UT باستخدام بيانات اختبار القدرة العددية في مقياس الخليج للقدرات العقلية، وبلغ حجم العينة 4206 طالبا من دول الخليج. ومن النتائج التي توصلت إليها هذه الدراسة إلى أن قيم مؤشر مربع البواقي (الموزون وغير الموزون) تتبع التوزيع الطبيعي بمتوسط حسابي لا يساوي الصفر، وهذا يتعارض مع ما توصل له (Almehrizi, 2010) في دراسته والتي أشار فيها إلى أنه إذا تطابقت البيانات مع النموذج المستخدم فإن كلا من المؤشرين WSR وUSRتتبع التوزيع الطبيعي بمتوسط حسابي صفر وانحراف معياري 1. وقد أرجعت الباحثة ذلك إلى اعتمادها على بيانات حقيقية بينما اعتمد المحرزي في دراسته على بيانات مولدة.

ISSN: 2706-6495

كما توصلت الدراسة إلى أن مؤشر مربع البواقي الموزون جاء في المرتبة الثالثة وغير الموزون جاء في المرتبة الرابعة في الكشف عن عدد الاستجابات غير المطابقة وذلك بعد مؤشري رايت الموزون ودرازجو في حين جاء مؤشر رايت غير الموزون في المرتبة الأخيرة. ومن النتائج التي توصلت لها الدراسة فيما يتعلق بصدق المؤشرات جاء مؤشر مربع البواقي الموزون ومؤشر درازجو كأكثر المؤشرات صدقا من بين المؤشرات الخمسة، وأظهرت الدراسة أن الغش كان أكثر الأنماط غير المطابقة انتشارا وأنه أكثر نسبة بين الذكور منه عند الإناث. وجميع هذه الدراسات أشارت إلى تمتع مؤشر مربع البواقي على مستوى المفردة بخصائص إحصائية جيدة.

أما بالنسبة لمؤشر مربع البواقي البينية، فيعتمد على كمية التطابق في نتائج الفرد على كل قسم من أقسام الاختبار وفحص إمكانية تفسير هذا التطابق وفق نموذج IRT المستخدم. ولهذا المؤشر أيضا نسختان: موزونة وغير موزونة (UBSR & WBSR). تستخدم النسختان لمؤشر مربع البواقي البينية تقدير مقدار التطابق $BSR_{\alpha j}$ في درجات الطالب α في كل قسم j من خلال حساب مربع فرق البواقي:

$$BSR_{\alpha j} = \left(X_{\alpha j} - \sum_{i=1}^{n_j} p_{\alpha i}\right)^2$$

حيث $X_{\alpha j}$ تمثل مجموع درجات الطالب α في فقرات القسم j و n_j يمثل عدد فقرات القسم j.

تأخذ $BSR_{\alpha j}$ أي قيمة مستمرة/متصلة بين الصفر و n_j^2 لأي طالب حسب مستويات قدرتهم. وللحكم على تقدير التطابق واستخدامه في تصنيف استجابات الطلبة على أنها ملائمة للنموذج أو غير ملائمة يتطلب معايرة من خلال حساب المتوسط الحسابي والانحراف المعياري لهذا التقدير من التطابق $BSR_{\alpha j}$. في ظل افتراض الاستقلال الموضعي لاستجابات الفقرات باستخدام IRT، فإن المتوسط الحسابي لـ $BSR_{\alpha j}$ لكل قسم هي قيمة التباين المعروف في IRT للدرجة الإجمالية لكل قسم بالاختبار بمعطى القدرة:

$$E(BSR_{\alpha j}) = \sum_{X_{j=0}}^{n_j} \left(X_j - \sum_{j=1}^{n_j} p_{aj}\right)^2 f(X_j|\Theta_a) = \sum_{i=1}^{n_j} p_{ai}q_{ai}$$

وبالمثل، يمكن الحصول على تباين فرق مربع البواقي، $BSR_{\alpha j}$ لكل قسم من خلال معادلة التباين النموذجية:

$$Var(BSR_{aj}) = \sum_{X_j=0}^{n_j} \left[\left(X_j - \sum_{i=1}^{n_j} p_{ai}\right)^2 - \sum_{i=1}^{n_j} p_{ai}q_{ai}\right]^2 f(X_j|\Theta_a)$$

$$= \sum_{X_j=0}^{n_j} \left(X_j - \sum_{i=1}^{n_j} p_{ai}\right)^4 f(X_j|\Theta_a) - \left[\sum_{i=1}^{n_j} p_{ai}q_{ai}\right]^2$$

حيث X_j هي جميع الدرجات الممكنة على كل مجموعة من المفردات j و $f(X_j|\Theta_a)$ هي دالة الكثافة الاحتمالية لدرجات X_j عند توفر قدرة الممتحن a . ويمكن الحصول على هذه الدالة من خلال الصيغة الدورية Lord and Wingersky (1984) recursion formula التي تم توظيفها من قبل المحرزي (Almehrizi, 2019).

للحصول على $f(X_j|\Theta_a)$ لكل قسم، يتطلب تعريف X_j كمتغير عشوائي للدرجات الخام على الفقرات الأولى في قسم الاختبار، وبالتالي تمثل $f(X_j = x_j|\Theta_a)$ دالة الكثافة الاحتمالية لكل X_j عندما تكون مساوية لـ X_j في قسم الاختبار j المكون

ISSN: 2706-6495

من i من الفقرات. ويتم تطبيق الصيغة الدورية لكل قسم من أقسام الاختبار ونبدأ بافتراض أنه مكون من فقرة واحدة، ويتم إدخال $i = 1$ في الصيغة:

$$f(X_{j1} = 0|\Theta_a) = 1 - p_{a1}$$

$$f(X_{j1} = 1|\Theta_a) = p_{a1}$$

ثم ندخل بقية الفقرات في ذلك القسم بحيث تكون الصيغة بالنسبة إلى $i > 1$ كما يلي:

$$f(X_{ji} = x_i|\Theta_a) = f(X_{j(i-1)} = x_j|\Theta_a)(1 - p_{ai})$$

$$+f(X_{ji} = x_j - 1|\Theta_a)p_{ai} \quad \text{for} \quad x_{ji} = 0,1, \ldots . i$$

لاستخدام الصيغة الدورية، ندخل الفقرات في الصيغة الدورية بأي ترتيب، بدءًا من $i = 1$ ، ونقوم بتطبيق الصيغة بشكل متكرر عن طريق زيادة i في كل تكرار، وتتوقف العملية بعد $i = n_j$ ، مما يعطي الصيغة المطلوبة $f(X_j|\Theta_a)$ حيث

$$f(X_j|\Theta_a) = f\left(X_{jn_j} = x_{n_j}|\Theta_a\right)$$

ويمكن الحصول على النسخة الموحدة القياسية لمربع فرق البواقي عبر جميع الأقسام (المؤشر غير الموزون) بطريقتين: تتطلب الطريقة الأولى توحيد BSR_{aj} في كل قسم، ثم جمع جميع الأقسام وأخيراً القسمة على الجذر التربيعي لعدد الأقسام (J)، على النحو التالي:

$$UBSR_a = \frac{1}{\sqrt{J}} \sum_{j=1}^{J} \frac{BSR_{aj} - E(BSR_{aj})}{\sqrt{Var(BSR_{aj})}}$$

والطريقة الأخرى لتشكيل النسخة الموحدة لمربع فرق البواقي عبر جميع الأقسام (المؤشر الموزون) من خلال تجميع مقدار التطابق في درجات الشخص عبر جميع مجموعات الفقرات ثم حساب النسخة الموحدة القياسية باستخدام الإستقلال الموضعي لأقسام الاختبار.

$$WBSR_a = \frac{BSR_a - E(BSR_a)}{\sqrt{Var(BSR_a)}}$$

حيث

$$BSR_a = \sum_{j=1}^{J} BSR_{aj}$$

$$E(BSR_a) = \sum_{j=1}^{J} E(BSR_{aj}) = \sum_{i=1}^{n} p_{ai} q_{ai}$$

$$Var(BSR_a) = \sum_{j=1}^{J} Var(BSR_{aj})$$

إذا كان نموذج IRT المستخدم يناسب بيانات الاختبار، فمن المفترض أن كلا المؤشرين (الموزون وغير الموزون) يتبع نظرية التوزيع الطبيعي الاعتدالي. وتعتبر الدرجات المتطرفة لمؤشر مطابقة الفرد على مستوى القسم في الطرف الأيمن من التوزيع الطبيعي الاعتدالي أنماطا غير مطابقة للاستجابة وهي القيم التي تزيد عن القيمة 1.645 عند مستوى دلالة 0.05 .(Almehrizi, 2019)

وقام المحرزي (Almehrizi, 2019) بالمقارنة بين أربعة مؤشرات لملائمة الفرد باستخدام البواقي على أقسام الاختبار باستخدام نماذج المفردة الثنائية وهما مؤشري مربع البواقي البينية الموزون وغير الموزون، ومؤشري (Wright, 1977) الموزون وغير الموزون. كما تحققت الدراسة من الخصائص الإحصائية لهذه المؤشرات، وذلك بتطبيقها على بيانات مولدة افتراضية (بالتحكم في عدد أقسام الاختبار وعدد الفقرات في كل قسم)، وأخرى مولدة من معالم حقيقية لاختبار القدرة اللفظية في مقياس الخليج للقدرات العقلية تتوزع فيه القدرات على مستويات محددة (3، 2، 1، 0، -1، -2، -3) بحيث تتبع التوزيع الطبيعي. وتوصلت الدراسة إلى امتلاك مؤشري مربع البواقي البينية خصائص إحصائية متميزة على عكس ما ظهر عند تطبيق مؤشري (Wright, 1977)، حيث كانت قيم المتوسط الحسابي تقترب جدا من الصفر لجميع القدرات ضمن جميع إجمالي البيانات حتى تلك البيانات التي تحتوي على أصغر عدد من الأقسام وأصغر عدد من المفردات داخل كل قسم، بينما في مؤشري رايت فقد كانت قيم متوسطه الحسابي منحرفة عن الصفر في جميع مستويات القدرة في إجمالي مجموعة البيانات. كما كانت قيم الانحراف المعياري في مؤشري مربع البواقي البينية تقترب من القيمة النظرية (الواحد) عند جميع القدرات، بينما انحرفت قيم الانحراف المعياري في مؤشري رايت عن الواحد، وتأثرت بعدد الأقسام أكثر من تأثرها بعدد المفردات لكل قسم. وبالنسبة لمعدلات الخطأ من النوع الأول Type I error فقد كانت متضخمة في مؤشري رايت في جميع مستويات القدرة في جميع البيانات وتأثرت معدلات الخطأ بعدد أقسام الاختبار حيث أدت زيادة عدد الأقسام إلى تحسين قدرة مؤشري رايت على التحكم في معدلات الخطأ من النوع الأول. وفي المقابل، كانت معدلات الخطأ من النوع الأول في مؤشري مربع البواقي البينية قريبة من 0.05 في مختلف مجموعات البيانات. وبالنسبة لقوة الكشف عن الاستجابات غير المتطابقة Power بالنسبة لمؤشر مربع البواقي البينية فقد كان قادرا على اكتشاف الاستجابات غير المطابقة حتى مع مستويات القدرة التي كان فيها لمؤشر رايت معدلات خطأ متضخمة من النوع الأول. كما أدت زيادة عدد المفردات لكل قسم إلى تحسين قوة هذا المؤشر.

1.1. مشكلة الدراسة:

تعتمد الدول المتقدمة اعتمادًا كبيرًا ورئيسيا في تطوير أنظمتها التعليميَّة على الدِراسَات والبحوث التربوية المتعددة، وتعتبر نتائج اختبارات TIMSS أحد المصادر التي يمكن الاعتماد عليها في تطوير أنظمة وممارسات التعليم في العلوم والرياضيات في الدول المشاركة. والمتتبع للأبحاث التي تتم على نتائج TIMSS في الدول المشاركة يلاحظ تركيزها على عوامل ومتغيرات متعددة، ومحاولتها التعرُّف على أثرها في تحصيل الطلبة في هاتين المادتين، وينبغي على دول الخليج ومنها السلطنة أن تولي هذه النتائج أهميَّة أكبر من حيث التحليل والبحث، وعدم الاكتفاء بالنظر إلى النتائج المعلنة؛ أو إعداد تقرير وصفي لترتيب الدولة، أو معرفة نسب الطلبة الذين أجابوا عن فقرات معينة من الاختبار، حيث إنَّ هذا النوع من التقارير لا يساعد على الاستفادة من نتائج الدراسة بالصورة العميقة التي يمكن أن تنعكس على السياسات التطويريَّة للنظام التعليمي، وممارساته المختلفة، وخاصة على مستوى المدرسة والقاعة الصفية؛ ولذا ينبغي إجراء العديد من الدراسات والأبحاث التحليليَّة؛ للنَّظر في نتائج TIMSS من زوايا متعددة ومن أجل الوصول إلى شواهد وأدلة علميَّة تقود إلى تطوير السياسات والممارسات التعليميَّة (الشمراني وآخرون، 2016).

وقد شاركت السلطنة اعتبارًا منذ عام 2007 في دورات اختبارات TIMSS مع العلم أن هذه الدورات يتم تنظيمها مرة كل أربع سنوات وقد حصلت السلطنة في نتائج اختبارات الرياضيات والعلوم للصف الثامن على نتائج متدنية خلال الدورات السابقة، حيث حصلت في اختبار الرياضيات على 372، 366، 403، 411 نقطة للسنوات 2007، 2011، 2015، 2019 على التوالي،

ISSN: 2706-6495

كما حصلت في اختبار العلوم على 423، 420، 455، و457 نقطة على التوالي، مع العلم أن المتوسط العام يبلغ 500 نقطة. وفي الرياضيات حصلت السلطنة على ترتيب 41 وفي العلوم على ترتيب 36 من 48 دولة مشاركة في عام 2007، وحصلت على ترتيب 41 في الرياضيات وفي العلوم على ترتيب 36 من 42 في دورة 2011، وعلى 32 في الرياضيات و 29 في العلوم من 39 في دورة 2015، وعلى ترتيب 35 في الرياضيات و 30 في العلوم من 39 في دورة 2019 (TIMSS & PIRLS, n.d)

وقد يرجع تدني هذه النتائج إلى كون نسبة من استجابات الطلبة غير مطابقة على أسئلة اختبارات TIMSSمما يدعونا إلى ضرورة دراسة هذه النتائج دراسة متأنية والتعرف على كونها ناتجة عن استجابات متوقعة حسب مستويات الطلبة أو عن استجابات غير مطابقة ناتجة عن عوامل مختلفة كالقلق والتخمين وغيرها، وذلك بتوظيف مؤشرات مطابقة الفرد التي ثبتت فعاليتها. ومن الجدير بالذكر أن اكتشاف أسباب وجود الاستجابات غير المطابقة في اختبارات دولية كاختبار TIMSS يساعد المختصين في مراجعة السياسات التعليمية وبالتالي التوصل إلى نتائج أفضل دوليا.

ومن خلال استعراض الدراسات السابقة والتي تناولت نتائج TIMSS نجد أنها ركزت على أسباب تدني نتائج الطلبة من حيث كونها راجعة لعوامل مختلفة كممارسات المعلمين واتجاهات الطلبة ذوي التحصيل المرتفع وغيرها إلا أن القليل منها ما اهتمت بموضوع التأكد من أن هذه النتائج تمثل استجابات الطلبة الحقيقية. ومن الدراسات التي تناولت نتائج TIMSS بالدراسة، دراسة عضيبات (2019) والتي هدفت لمقارنة ممارسات معلمي العلوم للصف الثامن في الأردن وسنغافورة في ضوء نتائج TIMSS2015 للكشف عن إسهام هذه الممارسات في تباين نتائج الطلبة بين الدولتين وتم الاستعانة باستبانة معلمي العلوم في الأردن وسنغافورة المشاركين في TIMSS2015 وكان من نتائج الدراسة وجود فروق دالة إحصائيا في الممارسات التدريسية والممارسات المتعلقة بالشعور بالثقة لدى المعلم أثناء التدريس لصالح سنغافورة. وهدفت دراسة العلوي (2017) للكشف عن مدى تضمن كتب العلوم للصفوف (5-8) بسلطنة عمان لموضوعات اختبارات TIMSS2015 حيث تم إعداد بطاقة تحليل المحتوى في ضوء هذه الموضوعات بعد أن تمت ترجمتها وعرضها على ذوي الاختصاص، وأظهرت النتائج أن محتوى كتب العلوم قد تضمن موضوعات اختبارات TIMSS2015 بنسب متفاوتة، وأوصت الدراسة بضرورة إطلاع مؤلفي كتب العلوم على موضوعات TIMSSوتضمينها بما يتفق مع النسب التي نصت عليها والاستفادة من الدول المشاركة والتي حققت مراكز متقدمة وكيف تضمنت كتب العلوم فيها لهذه الموضوعات.

كما هدفت دراسة شحادة والقراميطي (2016) للتعرف على أسباب تدني مستوى تحصيل طلبة المملكة العربية السعودية في نتائج TIMSS من وجهة نظر المعلمين والمشرفين من خلال إعداد استبيان مكون من 46 فقرة موزعة على أربعة مجالات: المنهج والمعلم والمتعلم والبيئة التعليمية. وقد توصلت الدراسة إلى عدة نتائج منها عدم جدية الطلبة في الإجابة عن أسئلة الاختبارات لأنها لا تدخل ضمن نتائجهم الدراسية. وسعت دراسة أبو عيش (2015) للتعرف على العادات الدراسية والخصائص الشخصية للطلبة الذين حصلوا على نتائج مرتفعة في اختبارات TIMSS عن غيرهم ممن حصلوا على نتائج منخفضة وقد اعتمدت الدراسة على نتائج طلبة المملكة العربية السعودية في TIMSS2003 وعلى الاستبيانات التي قاموا بتعبئتها كجزء من أدوات TIMSS2003 والتي تناولت الخلفية الأسرية للطلبة واتجاهاتهم وطموحاتهم والممارسات الصفية واللاصفية لمعلمي الرياضيات والعلوم واستخدامهم للحاسبات الآلية في داخل المدرسة وخارجها وتوصلت الدراسة إلى عدة نتائج منها أن اتجاهات الطلبة الحاصلين على نتائج مرتفعة نحو العلوم والرياضيات كانت أكثر إيجابية ممن حصلوا على نتائج متدنية.

ISSN: 2706-6495

ومن الدراسات التي اهتمت بالكشف عن استجابات الطلبة غير المطابقة في اختبارات TIMSS، دراسة الجراح (2020) والتي هدفت إلى التعرف على أنماط الاستجابة في اختبار TIMSS2015 في مبحثي العلوم والرياضيات لطلبة الصف الثامن في المملكة العربية السعودية وطلبة الصف الثامن في دولة سنغافورة باستخدام مؤشرات مطابقة الشخص التي طورها Huang, (2011) اعتمادا على مؤشري Within ability index (W) و Beyond ability index (B) التي طورها دا كوستا (D' Costa, 1993a, 1993b) و مؤشرات Huang هي (B_I^0 ،w_i^0 ،B_i^1 ،w_i^1) أو مؤشر التمكن Capability Index ومؤشر التخمين Guessing Index ومؤشر اللامبالاة Carelessness Index ومؤشر سوء الفهم Misconception Index وهذه المؤشرات تقوم عملية حسابها على مصفوفة جتمان (Guttman) القائمة على ترتيب قدرات المفحوصين من الأعلى إلى الأدنى وترتيب الفقرات من الأسهل إلى الأصعب. وأظهرت النتائج لتحليل بيانات دولة سنغافورة لمبحثي العلوم والرياضيات بأن 92% و94% على التوالي صنفت كأنماط استجابة طبيعية وأن8% و6% صنفت كأنماط استجابة غير طبيعية؛ بينما كانت نتائج التحليل للملكة العربية السعودية لمبحثي العلوم والرياضيات كانت 85% و86% على التوالي استجابات طبيعية و15% و14% غير طبيعية. وكانت أسباب تلك الاستجابات الشاذة للطلبة في كلتا الدولتين راجعة لعاملي التخمين واللامبالاة.

وتأتي أهمية مؤشرات مطابقة الفرد على مستوى أقسام الاختبار في أنها تساعد في الكشف عن الاستجابات غير المطابقة التي لا يمكن الكشف عنها باستخدام مؤشرات مطابقة الفرد على مستوى المفردة الواحدة، ومثال ذلك ما جاء في دراسة فيلت وآخرون (Felt et al., 2017)، والتي أظهرت أن استخدام مؤشرات المطابقة على مستوى المفردة أدى إلى عدم الكشف عن استجابات غير مطابقة لدى المستجيبين على أحد الاستبيانات المطبقة لأغراض دراسة صحية، بينما أظهرت مؤشرات المطابقة على مستوى أقسام الاختبار 18 استجابة غير مطابقة، والتي كان عدم الكشف عنها سيفوت الفهم والتعمق في المشكلات التي يعاني منها هؤلاء المستجيبون. ويعتبر مؤشر مربع البواقي البينية Between Residual Square Index من المؤشرات الحديثة والتي ظهرت مؤخرا على مستوى أقسام الاختبار، ويعتبر امتدادا للمقاربة التي قام بها المحرزي (2004؛ 2010) لمؤشر ملائمة الفرد على مستوى المفردة والمعتمد على البواقي حيث يعتمد المؤشر الجديد على كمية التطابق في نتائج الفرد على كل قسم من أقسام الاختبار وفحص إمكانية تفسير هذا التطابق وفق نموذج نظرية الاستجابة للمفردة IRT المستخدم أو أنه يشير إلى وجود نتائج شاذة لأقسام الاختبار (Almehrizi, 2019).

ويمكن تقسيم فقرات الاختبار حسب طبيعة السمة التي يقيسها من حيث مجالات المحتوى والمحاور المكونة للسمة النفسية أو السمة المعرفية. كما يمكن تقسيم فقرات الاختبار إلى أقسام وفقا للعمليات المعرفية التي يستهدفها الاختبار مثل تصنيف بلوم للمخرجات التعليمية في التحصيل الدراسي. وتساعد مؤشرات فحص الاستجابات المطابقة عبر أقسام الاختبار في الحكم على مدى مطابقة استجابات الطلبة بين مجالات المحتوى المختلفة أو العمليات المعرفية فمن المتوقع اختلاف الأداء الفعلي للطالب عن أدائه المتوقع في هذه الأقسام حسب مستوى قدرة الطالب. وفي المقابل يمكن توظيف نوع الفقرات المستخدمة في الاختبار ومستويات صعوبتها لتقسيم الاختبار وتوظيفها في مقارنة الأداء الفعلي للطالب مع الأداء المتوقع له في هذه الأقسام والكشف عن الطلبة ذوي الاستجابات غير المطابقة.

2.1. أسئلة الدراسة:

تستهدف الدراسة الحالية توظيف تنوع فقرات اختبار الرياضيات في TIMSS2019من حيث مجالات المحتوى التي تقيسها والعمليات المعرفية ونوع الفقرات المستخدمة ومستويات صعوبتها في الحكم على استجابات طلبة الصف الثامن بسلطنة عمان

وتحديد مدى انتشار الاستجابات الشاذة أو غير المطابقة باستخدام مؤشري مربع البواقي البينية. ويمكن تلخيص مشكلة الدراسة في الأسئلة التالية:

1- ما الإحصاءات الوصفية لقيم مؤشري مربع البواقي البينية (الموزون وغير الموزون) في اختبار الرياضيات للصف الثامن في TIMSS2019 بسلطنة عمان باستخدام أقسام الاختبار الأربعة: المحتوى، المستويات المعرفية، نوع الأسئلة، الصعوبة؟

2- ما نسبة الأفراد ذوي الاستجابات غير المطابقة في اختبار الرياضيات للصف الثامن في TIMSS2019بسلطنة عمان حسب مؤشري مربع البواقي البينية باستخدام أقسام الاختبار الأربعة: المحتوى، المستويات المعرفية، نوع الأسئلة، الصعوبة؟

3- هل تختلف نسبة الأفراد ذوي الاستجابات غير المطابقة في اختبار الرياضيات للصف الثامن في TIMSS2019بسلطنة عمان عند استخدام مؤشر مربع البواقي على مستوى المفردة وعند استخدام مؤشر مربع البواقي البينية على مستوى أقسام الاختبار الأربعة: المحتوى، المستويات المعرفية، نوع الأسئلة، الصعوبة؟

3.1. أهمية الدراسة:

تأتي أهمية هذه الدراسة من كونها أحد الدراسات القليلة التي تتناول الكشف عن الاستجابات غير المطابقة في نتائج الطلبة في اختبارات TIMSS مما قد يساعد المختصين في التقويم التربوي على اتخاذ قرارات أفضل لتحسين مستوى أداء الطلبة، حيث أنها تساعد في الكشف عن الاستجابات غير المطابقة للطلبة باستخدام مؤشرات مطابقة الفرد على مستوى المفردة وعلى مستوى أقسام الاختبار، والكشف عن وجود هذا النوع من الاستجابات ومعرفة نسبتها يساعد في تحديد أفضل لمشكلة تدني نتائج السلطنة في اختبارات TIMSS، فإذا كانت النسب مرتفعة فإن ذلك يستدعي البحث في الأسباب وراء هذه الاستجابات غير المطابقة للتعامل معها بما يقللها مستقبلا وبالتالي تحسن النتائج في الدورات القادمة لاختبارات TIMSS، أما إذا كانت نسب الاستجابات غير المطابقة متدنية، بمعنى أن النسبة الأعلى لاستجابات الطلبة طبيعية فإن ذلك يشير لوجود ضعف حقيقي لدى طلبة السلطنة وبالتالي ضرورة دراسة الأسباب المحتملة لهذا الضعف ووضع الخطط المناسبة لعلاج هذا الضعف.

وتأتي أهمية هذه الدراسة نظريا أيضا كونها تركز الاهتمام لمؤشرات مطابقة الفرد على مستوى أقسام الاختبار التي تم إهمالها بسبب ضعف خصائص التوزيع لمؤشرات مطابقة الفرد على مستوى أقسام الاختبار الموجودة سابقا، وذلك بإلقاء الضوء على أحد المؤشرات الحديثة المعتمدة على البواقي وهو مؤشر مربع البواقي البينية، حيث أنه من الضرورة بمكان عدم إهمال المؤشرات على مستوى أقسام الاختبار كونها تكشف الاستجابات غير المطابقة التي لا تتمكن المؤشرات على مستوى المفردة من الكشف عنها. وتعتبر هذه الدراسة أول دراسة يستخدم فيها مؤشر مربع البواقي البينية على بيانات ميدانية.

4.1. مصطلحات الدراسة:

مؤشر مطابقة الفرد: جاء في لوبيز ومونتيسونس (Lopez & Montesinos, 2005) وميجر وسيجيتيسما Sijtsma & Meijer (2001,) الذي ورد في حمادنة (2015) أنه مؤشر إحصائي يحدد المدى أو البُعد بين البيانات الفعلية المتمثلة باستجابات الأفراد والقيم المتوقعة من خلال النموذج المستخدم من خلال مقارنة قيم هذه المؤشرات مع قيمة حرجة لتحديد أنماط الاستجابة المطابقة وأنماط الاستجابة غير المطابقة.

مؤشر مربع البواقي البينية: مؤشر للكشف عن الاستجابات غير المطابقة للفرد على مستوى أقسام الاختبار، ويعتبر امتدادا لمؤشر مربع البواقي على مستوى المفردة الذي طوره المحرزي (Almehrizi, 2004; 2010) والذي يعتمد على منهج البواقي، ويعتمد مؤشر مربع البواقي البينية على مجموع درجات الفرد على كل قسم من أقسام الاختبار، وليس على درجات كل مفردة (Almehrizi, 2019).

5.1. حدود الدراسة:

- الحدود المكانية: تم تطبيق البحث على طلبة الصف الثامن الذين خضعوا لاختبار timss2010 في (سلطنة عمان)
- الحدود الزمنية: البحث يقتصر على نتائج طلبة سلطنة عمان في timss2019 المطبق عام 2019.
- تتحدد بيانات العينة المأخوذة من نتائج طلبة الصف الثامن فقط في اختبار الرياضيات في TIMSS2019.
- تقتصر الدراسة على استخدام النموذج ثلاثي المعالم المستند على مفاهيم نظرية الاستجابة للفقرة.

2. منهجية الدراسة والإجراءات

1.2. منهج الدراسة:

تعتمد الدراسة الحالية على المنهج الوصفي وذلك بوصف وتحليل البيانات التي تم جمعها من عينة الدراسة ومعالجتها في ضوء الأساليب الإحصائية؛ حيث تم تحليل بيانات طلبة الصف الثامن بسلطنة عمان في اختبار الرياضيات في TIMSS2019 باستخدام نظرية الاستجابة للمفردة، ثم حساب عدد ونسب الاستجابات غير المطابقة لهؤلاء الطلبة باستخدام مؤشري مربع البواقي البينية (Almehrizi, 2019) الموزون وغير الموزون عبر أقسام الاختبار، ومقارنتها بعدد ونسب الاستجابات غير المطابقة باستخدام مؤشري مربع البواقي (Almehrizi, 2010) الموزون وغير الموزون على مستوى المفردة.

2.2. مجتمع الدراسة والعينة:

استخدمت الدراسة بيانات أرشيفية لاستجابات طلبة الصف الثامن المشاركين في اختبارات الرياضيات في TIMSS2019 المطبقة عليهم في العام الدراسي (2019/2018) والبالغ عددهم عددهم (6745) واحتوت البيانات على بيانات 3341 طالبا و3404 طالبة موزعين على مختلف محافظات ومناطق سلطنة عمان.

3.2. أداة الدراسة:

استخدمت الدراسة اختبار الرياضيات في TIMSS 2019 والمكون من 14 كتيبا متكافئا، ويكون لكل طالب كتيب واحد لإنجازه. وتقوم الجمعية الدولية لتقويم التحصيل التربوي IEA لإعداد كتيبات اختبارات TIMSS باستخدام تقنية المصفوفة Matrix Sampling التي تقسم فيها الأسئلة إلى كتل أو مجاميع من الأسئلة بحيث يحتوي كل كتيب على مجموعتين من هذه الكتل أو مجاميع الأسئلة.

ويتراوح عدد الأسئلة في كل مجموعة من مجاميع الأسئلة بين 12 و18 سؤالا ويظهر كل سؤال في كتيبين، فعلى سبيل المثال يتكون الكتيب الأول من مجاميع الأسئلة التي تبدأ بالرمز MP01، ومجاميع الأسئلة التي تبدأ بالرمز MP02، والكتيب الثاني يتكون من مجاميع الأسئلة التي تبدأ بالرمز MP02، و MP03 وهكذا مما يوفر آلية لربط استجابات الطلبة من الكتيبات المختلفة معًا عند أخذ البيانات من جميع الكتيبات سويا (Mullis et al., 2017).

وتتوزع الأسئلة في اختبار الرياضيات على بعدين: بعد المحتوى والذي ينقسم إلى أربعة مجالات وهي الأعداد (30%) والجبر (30%) والهندسة (20%) والبيانات والاحتمالات (20%) وبعد العمليات المعرفية ويتكون من ثلاثة مستويات: المعرفة (35%) والتطبيق (40%) والاستدلال (25%) (Mullis et al., 2017) . وتأتي أسئلة اختبار الرياضيات في TIMSS2019 على نوعين: ثنائية الاستجابة وهي أسئلة اختيار من متعدد وأسئلة ذات إجابات قصيرة، والنوع الثاني متعددة الاستجابة (هيئة تقويم التعليم والتدريب، 2019).

وفي هذه الدراسة تم الاعتماد على الأسئلة ثنائية الاستجابة نظرا لكون مؤشر مربع البواقي البينية يتناسب مع هذا النوع من الاستجابات ولا يمكن تطبيقه على الأسئلة متعددة الاستجابة. وتم تقسيم الأسئلة ثنائية الاستجابة إلى قسمين حسب نوع الأسئلة: اختيار من متعدد والسؤال القصير. وتم تحويل الدرجات في أسئلة الاختيار من متعدد والأسئلة ذات الاستجابة الثنائية إلى صفر وواحد. كما تم تقسيم أسئلة الاختبار إلى ثلاثة أقسام حسب مستوى صعوبتها باستخدام معلمة الصعوبة في النموذج ثلاثي المعالم: (0-0.75) و (أكبر من 0.75) و(أقل من 0)، وتم اعتماد هذا التقسيم نظرا لأن أغلب الأسئلة في اختبار الرياضيات TIMSS2019 تتراوح بين الصعبة جدا إلى متوسطة الصعوبة. وتم توزيع الأسئلة ثنائية الاستجابة على المجالات في الأقسام المختلفة للاختبار (المحتوى، العمليات المعرفية، نوع الأسئلة، والصعوبة) من أجل إدخالها في برنامج R للقيام بالتحليلات المطلوبة، وكان توزيعها كما في جدول (1) والذي يوضح مجالات الاختبار وعدد الأسئلة بها وفقا للتقسيمات الأربعة في كل كتيب من كتيبات اختبار الرياضيات:

جدول (1) عدد الأسئلة موزعة على مجالات أقسام الاختبار في كتيبات الرياضيات مع عدد الطلبة لكل كتيب

الأقسام	الأقسام	الكتيبات														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	المجموع
عدد الطلبة		486	474	488	480	476	476	483	479	478	482	489	486	483	485	
إجمالي الأسئلة		30	28	27	28	26	26	30	27	26	27	27	26	29	33	390
المحتوى	أعداد	8	7	8	10	10	9	9	8	8	7	8	9	9	10	120
	جبر	10	9	9	9	6	5	9	10	9	9	9	8	8	10	120
	هندسة	6	6	4	3	4	5	5	5	6	6	5	5	7	7	74
	احتمالات	6	6	6	6	6	7	7	4	3	5	5	4	5	6	76
العمليات المعرفية	معرفة	7	9	9	11	10	9	10	8	7	6	9	10	10	9	124
	تطبيق	15	12	12	14	14	13	15	13	11	12	12	13	14	16	186
	استدلال	8	7	6	3	2	4	5	6	8	9	6	3	5	8	80
نوع الأسئلة	اختياري	15	12	12	12	15	10	14	11	10	11	13	12	11	16	174
	سؤال قصير	15	16	15	16	11	16	16	16	16	16	14	14	18	17	216

63	5	2	1	2	6	5	2	5	6	5	6	6	6	6	0 >	
202	18	18	17	17	12	11	15	16	12	13	13	11	16	16	0.75-0	الصعوبة
125	10	9	8	8	9	10	10	9	8	8	9	10	8	8	0.75<	

ويظهر من جدول (1) أن عدد الأسئلة في مجالي الأعداد والجبر يشكلان العدد الأكبر من الأسئلة في أقسام المحتوى، يليه مجال الاحتمالات ثم الهندسة. وبالنسبة لأقسام العمليات المعرفية يوجد العدد الأكبر من الأسئلة في مجال التطبيق يليه مجال المعرفة وأخيرا الاستدلال. وفي أقسام نوع الأسئلة فإن عدد الأسئلة القصيرة أكبر من عدد أسئلة الاختيار من متعدد. ويلاحظ من جدول (1) أيضا أن عدد الأسئلة أكبر في التقسيم (0-0.75) ثم (> 0.75) وأخيرا (أقل من 0).

4.2. الخصائص السيكومترية لاختبار الرياضيات في TIMSS2019

الصدق :

تم التحقق من مؤشرات الصدق لنتائج اختبار الرياضيات في TIMSS2019 من خلال حساب العلاقة الارتباطية باستخدام معامل ارتباط بيرسون بين درجات الطلبة على مفردات كل بعد وبين الدرجة الكلية لهذا البعد (الأعداد، الجبر، الهندسة، الاحتمالات)، وتراوحت معاملات الارتباط في جميع الكتيبات بين 0.14 و0.72 وكانت دالة إحصائيا عند مستوى 0.01 مما يدل على الاتساق الداخلي للاختبار.

الثبات:

وتم حساب مؤشرات الثبات لكل كتيب عن طريق معامل ثبات ألفا كرونباخ، وبطريقة التجزئة النصفية باستخدام برنامج SPSS، وعن طريق مؤشر بيتا (Almehrizi, 2021) باستخدام برمجية في لغة R باسم Galpha_Beta من إعداد الباحثين، وتراوحت قيم معاملات ثبات ألفا بين 0.80 و 0.88، وبين 0.76 و 0.88 باستخدام التجزئة النصفية، وبين 0.76 و 0.86 باستخدام مؤشر بيتا، وهي قيم ثبات مرتفعة.

5.2. التحقق من افتراضات نظرية الاستجابة للمفردة:

تم التحقق من افتراضات نظرية الاستجابة للمفردة في نتائج اختبار الرياضيات في TIMSS2019 على عينة الدراسة، وكانت النتائج كالتالي:

أولا: افتراض أحادية البعد

تم التحقق من افتراض أحادية البعد لكل كتيب في مادة الرياضيات باستخدام التحليل العاملي الاستكشافي عن طريق برنامج SPSS، واتضح أن التباين الكلي الذي يفسره العامل الأول في جميع الكتيبات تراوحت نسبته ما بين (18.51% و24.50%) وهي نسب أعلى من بقية العوامل، كما أن قيمة الجذر الكامن للعامل الأول في جميع الكتيبات مرتفعة عن بقية العوامل، وبقسمة الجذر الكامن للعامل الأول على الجذر الكامن للعامل الثاني كان الناتج أكبر من 2 في جميع الكتيبات، وهذا مؤشر على أحادية البعد كما ذكر جلورفيلد (Glorfeld,1995) ونستدل من ذلك على تحقق افتراض أحادية البعد.

ثانيا: الاستقلال الموضعي

تفترض نظرية الاستجابة للفقرة أن فقرات الاختبار مستقلة موضعيا بمعنى أن استجابة المفحوص لفقرة ما لا تتأثر إيجابا أو سلبا باستجابته على فقرة أخرى وإذا ما تم انتهاك افتراض الاستقلال الموضعي بين الفقرات سيظهر ما يسمى بالارتباط الموضعي

بين الفقرات واقترحت ين (Yen,1984) مؤشر Q3 للكشف عن الارتباط الموضعي بين الفقرات، ويعبر عنه بالعلاقة بين البواقي لزوج من الفقرات بعد ضبط السمة المقدرة. ويمكن الكشف عن الارتباط الموضعي للفقرات من خلال حساب الارتباطات بين بواقي أي فقرتين عند مستوى قدرة معين، فإذا كانت هذه العلاقة أكبر من الصفر دل ذلك على وجود الارتباط الموضعي بين الفقرات، وبالتالي فإن اقتراب قيم هذه الارتباطات من الصفر يعد إشارة إلى تحقق الاستقلال الموضعي (النعيمي، 2006). وتراوحت قيم معاملات الارتباط في جميع الكتيبات بين 0.03 و0.12، وتراوحت قيم متوسط مؤشر Q3 عند كل كتيب من كتيبات الاختبار بين 0.05 و0.07، وهي قيم صغيرة جدا لذلك نستطيع القول أن الأسئلة في كتيبات الرياضيات قد تحقق فيها افتراض الاستقلال الموضعي.

ثالثا: ملائمة النموذج

وبالنسبة لملائمة النموذج ثلاثي المعالم لمجموعة الأسئلة في الكتيبات فقد تم التحقق من ذلك باستخدام مؤشر البواقي المعيارية، حيث تم تقدير القدرات ببرنامج multilog وتم حساب البواقي ومؤشر البواقي المعيارية ومقارنتها مع مربع كاي كمؤشر لحسن مطابقة المفردات للنموذج، ويوضح جدول (2) المؤشر العام لمطابقة كل كتيب مع النموذج ثلاثي المعالم. ويتضح من جدول (2) تحقق ملائمة النموذج الثلاثي مع أسئلة الكتيبات حيث كانت قيم مؤشر البواقي المعيارية غير دالة مع قيمة مربع كاي والتي تساوي 3.84 (عند مستوى دلالة 0.05 ودرجة حرية 1).

جدول (2) المؤشر العام لمطابقة النموذج لمفردات اختبارات الرياضيات في TIMSS2019

رقم الكتيب	المؤشر العام	رقم الكتيب	المؤشر العام
1	0.76	8	0.67
2	0.78	9	0.69
3	0.79	10	0.77
4	0.78	11	0.81
5	0.85	12	0.77
6	0.89	13	0.76
7	0.73	14	0.78

6.2. تقديرات معالم مفردات اختبار الرياضيات في TIMSS2019

استخدمت الدراسة معالم مفردات اختبار الرياضيات في TIMSS2019 (الصعوبة والتمييز والتخمين) التي تم تقديرها من قبل الجمعية الدولية لتقويم التحصيل التربوي IEA، وتراوحت قيم معالم التمييز للمفردات بين 0.54 كأقل قيمة وكانت في المفردتين 21، و7 في الكتيبين العاشر، والحادي عشر على التوالي، وبين 2.29 كأعلى قيمة وكانت في المفردة 18 في الكتيب 11. أما بالنسبة لمعالم الصعوبة فتراوحت القيم بين 0.84- كأدنى قيمة وكانت في المفردتين 16، و 1 في الكتيبين 8، و 9 على التوالي، وبين 2.10 كأعلى قيمة وكانت في المفردتين 27، و 14 في الكتيبين 4، و 5 على التوالي. أما عن التخمين فتراوحت قيمة بين 0 كأدنى قيمة في 218 (55.90%) من المفردات، و0.48 كأعلى قيمة وكانت في المفردتين 24، و11 في الكتيبين 11، و12 على التوالي.

ISSN: 2706-6495

7.2. المعالجة الإحصائية:

للإجابة على السؤال الأول: ما الإحصاءات الوصفية لمؤشري مربع البواقي البينية الموزون وغير الموزون في اختبار الرياضيات للصف الثامن في TIMSS2019 بسلطنة عمان باستخدام أقسام الاختبار الأربعة: المحتوى، العمليات المعرفية، نوع الأسئلة، الصعوبة؟، تم حساب الإحصاءات الوصفية لمؤشري مربع البواقي البينية الموزون (WBSR) وغير الموزون (UBSR) من خلال برنامج R باستخدام برمجية باسم Section SRes من إعداد الباحثين، بحيث يتم إدخال ملف الاستجابات وملف معالم الأسئلة لكل كتيب في كل مرة لحساب قيم المؤشرات وإحصاءاتها الوصفية. المتوسط الموزون عبر كتيبات اختبار الرياضيات TIMSS2019 (لمراعاة اختلاف عدد فقرات الكتيبات) لكل من المتوسطات الحسابية والانحرافات المعيارية للمؤشرين.

وللإجابة عن السؤال الثاني: ما نسبة الأفراد ذوي الاستجابة غير المطابقة في اختبار الرياضيات للصف الثامن في TIMSS2019بسلطنة عمان حسب مؤشري مربع البواقي البينية باستخدام أقسام الاختبار الأربعة: المحتوى، العمليات المعرفية، نوع الأسئلة، الصعوبة؟ فقد تمت الإجابة عنه من خلال ثلاثة تحليلات إحصائية:

التحليل الأول: حساب عدد ونسبة الأفراد ذوي الاستجابات غير المطابقة في اختبار الرياضيات TIMSS2019 لكل مؤشر.

التحليل الثاني: مقارنة عدد ونسبة الأفراد ذوي الاستجابات غير المطابقة المتكررة وفقا لعدد أقسام الاختبار لكل مؤشر.

التحليل الثالث: دراسة توزيع تقدير القدرات للأفراد ذوي الاستجابات غير المطابقة لكل مؤشر.

وبالنسبة للتحليل الأول تم حساب عدد ونسبة الأفراد ذوي الاستجابات غير المطابقة وفق مؤشري مربع البواقي البينية الموزون WBSR وغير الموزون UBSR في الأقسام الأربعة في الاختبار، وتعتبر استجابات الفرد غير مطابقة إذا كانت قيمة المؤشر لها أكبر من القيمة 1.645 وهي الدرجة الحرجة المقابلة لمستوى الدلالة 0.05 في التوزيع الاعتدالي أحادي الذيل، حيث أن مؤشر مربع البواقي البينية يتبع التوزيع الاعتدالي.

وللإجابة عن السؤال الثالث: هل تختلف نسبة الأفراد ذوي الاستجابة غير المطابقة في اختبار الرياضيات للصف الثامن في TIMSS2019 بسلطنة عمان عند استخدام مؤشر مربع البواقي على مستوى المفردة وعند استخدام مؤشر مربع البواقي البينية على مستوى أقسام الاختبار الأربعة: المحتوى، المستويات المعرفية، نوع الأسئلة، الصعوبة؟ تم استخدام برمجية في لغة R باسم Item SRes من إعداد الباحثين بحيث يتم إدخال ملف الاستجابات في كل مرة لحساب قيمة المؤشرات لحساب عدد الأفراد ذوي الاستجابات غير المطابقة حسب مؤشري مربع البواقي على مستوى المفردة، ثم حساب نسبتها بالنسبة للعدد الكلي للاستجابات في كل كتيب، وتصنف الاستجابة على أنها غير مطابقة إذا كانت قيمة المؤشر أكبر من 1.645 وهي القيمة الحرجة المقابلة لمستوى الدلالة 0.05 في التوزيع الاعتدالي أحادي الذيل.

3. نتائج الدراسة

السؤال الأول: ما الإحصاءات الوصفية لمؤشري مربع البواقي البينية الموزون وغير الموزون في اختبار الرياضيات للصف الثامن في TIMSS2019 بسلطنة عمان باستخدام أقسام الاختبار الأربعة: المحتوى، العمليات المعرفية، نوع الأسئلة، الصعوبة؟

ISSN: 2706-6495

يوضح جدول (3) قيم المتوسط الموزون عبر كتيبات اختبار الرياضيات TIMSS2019 (لمراعاة اختلاف عدد فقرات الكتيبات) لكل من المتوسطات الحسابية والانحرافات المعيارية لمؤشري مربع البواقي البينية الموزون وغير الموزون. وبالنظر للمتوسط العام للمتوسطات الحسابية في جميع كتيبات اختبار الرياضيات في TIMSS2019 يظهر أنها تراوحت بين -0.30 و -0.23 وبين -0.29 و-0.24 لمؤشري UBSR وWBSR على الترتيب، وأن أقل قيمة للمتوسط الموزون كانت في أقسام العمليات المعرفية، حيث بلغت -0.3 لمؤشر UBSR و -0.29 لمؤشر WBSR، بينما كانت أكبر قيمة في أقسام نوع الأسئلة حيث بلغت -0.23 لمؤشر UBSR و -0.24 لمؤشر WBSR.

جدول (3) الإحصاءات الوصفية لمؤشري مربع البواقي البينية الموزون وغير الموزون لإجمالي كتيبات اختبار الرياضيات

الأقسام	UBSR		WBSR	
	المتوسط الحسابي	الانحراف المعياري	المتوسط الحسابي	الانحراف المعياري
المحتوى	0.27-	0.83	0.25-	0.84
العمليات المعرفية	0.30-	0.77	0.29-	0.76
نوع الأسئلة	0.23-	0.75	0.24-	0.78
الصعوبة	0.26-	0.78	0.26-	0.73

أما بالنسبة للمتوسط العام للانحرافات المعيارية في جميع كتيبات اختبار الرياضيات في TIMSS2019 للمؤشرين UBSR وWBSR على الترتيب فتراوحت بين 0.75 و0.83 وبين 0.73 و0.84، وأكبر قيمة كانت في أقسام المحتوى حيث بلغت 0.83 لمؤشر UBSR و 0.84 لمؤشر WBSR، بينما كانت أقل قيمة في أقسام نوع الأسئلة بالنسبة لمؤشر UBSR حيث بلغت 0.75 وفي أقسام الصعوبة بالنسبة لمؤشر WBSR حيث بلغت 0.73.

السؤال الثاني: ما نسبة الأفراد ذوي الاستجابة غير المطابقة في اختبار الرياضيات للصف الثامن في TIMSS2019بسلطنة عمان حسب مؤشري مربع البواقي البينية باستخدام أقسام الاختبار الأربعة: المحتوى، العمليات المعرفية، نوع الأسئلة، الصعوبة؟

يظهر جدول (4) إجمالي عدد ونسبة الأفراد ذوي الاستجابات غير المطابقة وفقا لمؤشري مربع البواقي البينية الموزون وغير الموزون لأقسام اختبار الرياضيات TIMSS2019. وتظهر النتائج أن تعدد الأقسام ساعد في الكشف عن عدد أكبر من الأفراد الذين يظهرون استجابات غير مطابقة. وبالنظر لإجمالي عدد الأفراد ذوي الاستجابة غير المطابقة في جميع الكتيبات يظهر أن عدد الأفراد ذوي الاستجابات غير المطابقة في أقسام نوع الأسئلة أكبر من بقية الأقسام، يليه أقسام المحتوى، ثم أقسام الصعوبة وأخيرا أقسام العمليات المعرفية.

جدول (4) إجمالي عدد ونسبة الأفراد ذوي الاستجابات غير المطابقة وفقا لمؤشري مربع البواقي البينية الموزون وغير الموزون لأقسام اختبار الرياضيات TIMSS2019

الأقسام	UBSR	WBSR
المحتوى	(%3.16) 213	(%3.28) 221
العمليات المعرفية	(%2.16)146	(%2.13) 144

نوع الأسئلة	(%3.22) 217	(%3.88) 262
الصعوبة	(%2.42) 163	(%2.51) 169

أما فيما يتعلق بالتحليل الثاني حول حساب عدد ونسبة الأفراد ذوي الاستجابات غير المطابقة المتكررة وفقا لعدد أقسام الاختبار التي تم فيها الكشف عن عدم مطابقة استجاباتهم، اتضح أن النسبة الأكبر من الطلبة تم الكشف عن استجاباتهم غير المطابقة في قسم واحد فقط من الأقسام الأربعة (المحتوى أو العمليات المعرفية أو نوع الأسئلة أو الصعوبة) دون بقية الأقسام، بمعنى أن الأفراد الذين يظهرون استجابات غير مطابقة وفقا لأقسام المحتوى مثلا يختلفون عن أولئك الذين يظهرون استجابات غير مطابقة وفقا للأقسام الأخرى، وكذلك الذين يظهرون استجابات غير مطابقة وفقا لأقسام العمليات المعرفية كذلك يختلفون عن أولئك الذين يظهرون استجابات غير مطابقة وفقا للأقسام الأخرى، وهكذا لبقية الأقسام. ويمثل عدد هؤلاء الأفراد 651 طالبا وبنسبة 94% من إجمالي الطلبة ذوي الاستجابات غير المطابقة حسب مؤشرUBSR، و652 طالبا بنسبة 90% حسب مؤشر WBSR. أما الطلبة الذين تظهر استجاباتهم غير مطابقة وفقا لقسمين من الأقسام الأربعة فيمثلون 44 طالبا وبنسبة 6% حسب مؤشر UBSR و59 طالبا بنسبة 8% حسب مؤشر WBSR. وأظهرت النتائج أيضا عدم وجود أفراد ذوي استجابات غير مطابقة تكرروا في ثلاثة أو أربعة أقسام حسب مؤشر UBSR، بينما ظهر 12 طالبا وبنسبة 5% في ثلاثة أقسام حسب مؤشر WBSR.

أما بالنسبة للتحليل الثالث والمتعلق بمستويات قدرات الطلبة الذين يظهرون استجابات غير مطابقة، فأظهرت النتائج أن قدراتهم تراوحت بين -2.04 و1.64 حسب المؤشرين، وتوزعت القدرات للأفراد ذوي الاستجابات غير المطابقة حسب المؤشر UBSR بشكل أكبر في المستويات بين (-1، 1) في أقسام المحتوى والعمليات المعرفية والصعوبة، بينما توزعت بشكل أكبر في المستويات (أقل من -1) في أقسام نوع الأسئلة في أغلب الكتيبات. بينما توزعت القدرات للأفراد ذوي الاستجابات غير المطابقة حسب المؤشر WBSR بشكل أكبر في المستويات بين (-1، 1) في أقسام المحتوى والعمليات المعرفية في أغلب الكتيبات، بينما توزعت بشكل أكبر في المستويات (أقل من -1) في أقسام نوع الأسئلة، بينما توزعت على المستويات بين (-1،1) و (أقل من -1) في أقسام الصعوبة. وأظهرت النتائج أن القدرات المنخفضة (الأقل من -1) تراوحت نسبتها بين 16% و54% وفق مؤشر UBSR عبر الكتيبات، وبين 17% و 65% وفق مؤشر WBSR. أما القدرات المتوسطة (بين-1 و 1) فتراوحت بين 44% و82% وفق مؤشر UBSR، وبين 33% و81% وفق مؤشر WBSR. وبخصوص القدرات المرتفعة (أكبر من 1+) فقد تراوحت بين 2% و5% وفقا لمؤشر UBSR، وبين 2% و4% وفقا لمؤشر WBSR. وتدل هذه النتائج على أن الأفراد ذوي الاستجابات غير المطابقة وفقا للمؤشرين هم من القدرات المنخفضة والمتوسطة ونسبة صغيرة من الطلبة ذوي القدرات المرتفعة ممن أظهروا استجابات غير مطابقة.

السؤال الثالث: هل تختلف نسبة الأفراد ذوي الاستجابة غير المطابقة في اختبار الرياضيات للصف الثامن في TIMSS2019 بسلطنة عمان عند استخدام مؤشر مربع البواقي على مستوى المفردة وعند استخدام مؤشر مربع البواقي البينية على مستوى أقسام الاختبار الأربعة: المحتوى، المستويات المعرفية، نوع الأسئلة، الصعوبة؟

يظهر جدول (5) ملخص لعدد ونسبة الأفراد ذوي الاستجابات غير المطابقة وفقا للمؤشرين USR و WSR في كتيبات اختبار الرياضيات TIMSS2019. وتظهر النتائج أن عدد الاستجابات غير المطابقة وفقا لمؤشري مربع البواقي تشكل نسبة منخفضة جدا من مجموع الاستجابات الكلية،

حيث بلغ عدد الطلبة 20 طالبا بنسبة 0.30% فقط وفقا لمؤشر USR، و64 طالبا 0.95% وفقا لمؤشر WSR، وتظهر النتائج عدم وجود أفراد ذوي استجابات غير مطابقة في بعض الكتيبات وفقا للمؤشرين USR و WSR (كتيب 10 في WSR والكتيبات 9، 10، 13 في USR).

جدول (5) عدد ونسبة الأفراد ذوي الاستجابات غير المطابقة وفقا للمؤشرين USR و WSR في كتيبات اختبار الرياضيات TIMSS2019

الملخص	USR		WSR	
	العدد والنسبة	الكتيبات	العدد والنسبة	الكتيبات
أصغر عدد للطلبة	(% 0) 0	9،10،13	(% 0) 0	10
أكبر عدد للطلبة	(%0.82) 4	4	(%2.88) 14	12
إجمالي الطلبة	(% 0.30) 20	-	(% 0.95) 64	-

وتبدو نسب الأفراد ذوي الاستجابات غير المطابقة وفق مؤشري مربع البواقي على مستوى المفردة WSR وUSR صغيرة مقارنة بنتائج مؤشري مربع البواقي البينية على مستوى أقسام الاختبار UBSR وWBSR ، حيث تراوحت النسب للاستجابات غير المطابقة بين 2.16% و3.22%حسب مؤشر UBSR، وبين 2.13 % و 3.88% حسب مؤشر WBSR؛ مما يعني أن هناك اختلافا بين الطريقتين في نسب الطلبة ذوي الاستجابات غير المطابقة في الاختبار.

وبدراسة توزيع قدرات الطلبة الذين يبدون استجابات غير مطابقة وفقا للمؤشرين USR وWSR أظهرت النتائج أنها قدرات منخفضة تتراوح بين -4 و-2.10 وفقا لمؤشر USR، وبين -4 و-1.94 وفقا للمؤشر WSR، مع وجود عدد قليل من الأفراد ذوي القدرات المتوسطة التي تتراوح بين -0.24 و0.94 حسب المؤشر USR وعددهم 8 أفراد، وبين -0.62 و0.24 حسب المؤشر WSR، وعددهم 4 طلبة، ولا يوجد حسب المؤشرين أي أفراد ذوي قدرات مرتفعة (أكبر من 1). وأظهرت النتائج أن النسبة الأعلى من القدرات للطلبة الذين يظهرون استجابات غير مطابقة كانت من القدرات المنخفضة (أقل من 1-)، وتمثل 70% من القدرات وفق مؤشر USR، و97% حسب مؤشر WSR، أما القدرات المتوسطة (بين 1- و 1+) فبلغت نسبتها 30% حسب مؤشر USR، و3% فقط حسب مؤشر WSR، ولا توجد قدرات عليا من بين هذه القدرات وفقا لكلا المؤشرين.

ويمكن القول أن مؤشري مربع البواقي و مربع البواقي البينية اتفقا في كون الاستجابات غير المطابقة في اختبار الرياضيات في TIMSS2019 كانت لدى الطلبة ذوي القدرات المنخفضة والمتوسطة، ولكنهما اختلفا في مدى هذه القدرات حيث نجد أن أدنى قدرة لدى مؤشر مربع البواقي البينية كانت 2.04- وأكبرها كانت 1.67 بينما كانت أدنى قدرة لدى مؤشر مربع البواقي -4 وأكبر قدرة 0.94 .

4. مناقشة النتائج

هدفت الدراسة إلى التعرف على درجة انتشار الاستجابات غير المطابقة أو الشاذة لدى طلبة الصف الثامن بسلطنة عمان في اختبار الرياضيات TIMSS2019 من خلال تطبيق مؤشري مربع البواقي البينية عبر أقسام الاختبار المرتبطة بالبناء المعرفي للاختبار (مجالات المحتوى ومستويات العمليات المعرفية) والأقسام المرتبطة بأسئلة الاختبار (نوع الأسئلة وصعوبتها).

وأظهرت نتائج الدراسة أن قيم المتوسط الحسابي لمؤشري مربع البواقي البينية الموزون وغير الموزون تنحرف قليلا عن الصفر، وأن الانحراف المعياري لها ينحرف عن الواحد، وبينما لا تتفق هذه النتائج مع نتائج دراسة المحرزي (Almehrizi, 2019)، حيث كانت قيم المتوسط الحسابي للمؤشرين تقترب جدا من الصفر لجميع القدرات ضمن جميع البيانات، وقيم الانحراف المعياري تقترب من القيمة النظرية (الواحد) عند جميع القدرات، إلا أن ذلك لا يعد اختلافا لكون دراسة (Almehrizi, 2019) طبق فيها المؤشران على بيانات مولدة افتراضية وأخرى مولدة من معالم ميدانية لاختبار القدرة اللفظية تتوزع فيه القدرات على مستويات محددة (3، 2، 1، 0، -1، -2، -3) بحيث تتبع التوزيع الطبيعي، بينما استخدمت هذه الدراسة بيانات ميدانية، حيث أن قدرات الطلبة في اختبار الرياضيات في TIMSS2019 تتوزع على مدى واسع من القدرات للطلبة في الكتيبات الأربعة عشر والتي تتوزع توزيعا ملتويا التواء موجبا أي تغلب القدرات المنخفضة لدى الطلبة.

أظهرت نتائج الدراسة أن عدد الأفراد ذوي الاستجابات غير المطابقة من مجموع عينة سلطنة عمان يختلف باختلاف أقسام اختبار الرياضيات في TIMSS2019، وأظهرت النتائج أيضا أن الطلبة الذين يظهرون استجابات غير مطابقة يتوزعون على الأقسام المختلفة للاختبار، ولا يشتركون في جميع الأقسام وإنما يشتركون بنسبة قليلة في قسمين وبنسبة أقل جدا في ثلاثة أقسام. ويمكننا الاستدلال من ذلك على أهمية تقسيم الاختبار إلى عدة أقسام، حيث أن تنوع الأقسام كشف عن حالات لم يكن من الممكن الكشف عنها في حال تم الاعتماد على قسم واحد.

وأظهرت النتائج أن عدد الأفراد غير المطابقين أكبر ما يمكن في أقسام نوع الأسئلة يليه في أقسام المحتوى، ثم أقسام الصعوبة وأخيرا أقسام العمليات المعرفية. ويمكن أن نفسر سبب كون نسبة الاستجابات غير المطابقة أعلى في أقسام نوع الأسئلة (أسئلة اختيار من متعدد وأسئلة الإجابات القصيرة) بسبب سلوك التخمين الذي قد يلجأ إليه الطلبة في أسئلة الاختيار من متعدد مقارنة بأسئلة الإجابة القصيرة، وقد أشار سوناس وآخرون (Iasonas et al., 2000) إلى أن التخمين من العوامل الرئيسية في ظهور أنماط الاستجابات غير المطابقة (بني عطا، 2019). ويمكن أيضا بسبب سلوك الغش حيث أن نقل الطالب للإجابة من زميله أسهل في حالة الأسئلة الموضوعية ومنها أسئلة الاختيار من متعدد، كما أن سلوك الغش في الأسئلة الموضوعية لا يتطلب أن يكون هناك اتصالا بصريا بورقة الزميل المجاور بل يمكن أن يتم عن طريق إشارات معينة (Belov et al., 2010).

وأظهرت النتائج أن قيم المؤشرات ومتوسطاتها الحسابية في جميع الكتيبات تتجه ناحية القيم المنخفضة مما يشير إلى أن أغلب الاستجابات في اختبار الرياضيات في TIMSS2019 طبيعية، ويدل على ذلك انخفاض نسب الأفراد ذوي الاستجابات غير المطابقة عبر جميع الكتيبات والتي تراوحت بين 2.16% و3.22%حسب مؤشر UBSR، وبين2.13 % و 3.88% من مجموع عينة سلطنة عمان حسب مؤشر WBSR.

وبالنظر في قدرات الطلبة الذين يظهرون استجابات غير مطابقة تم التوصل إلى أن أغلبها قدرات متوسطة و منخفضة، وربما يعود ذلك إلى أن العدد الأكبر من الأسئلة كان من مستوي التطبيق والاستدلال أي من أسئلة القدرات العليا التي يلجأ فيها الطالب متدني المستوى عادة إلى التخمين، كما يلاحظ أيضا أن الطلبة ذوي القدرات منخفضة أظهروا استجابات غير مطابقة أكثر في أقسام نوع الأسئلة (أسئلة اختيار من متعدد، وأسئلة ذات إجابات قصيرة)، والتي يكون فيه احتمال لجوء الطالب إلى التخمين أو الغش مرتفعة خصوصا إذا كان متدني التحصيل، ويتفق ذلك مع نتائج دراسة تشن (Chen, 2004) التي أشارت إلى أن الطلبة منخفضي التحصيل يلجأون للتخمين أكثر من الطلبة مرتفعي التحصيل في اختبار لغة انجليزية طبق على طلبة صينين. كما تتفق هذه النتائج مع دراسة براون وفيلاريل (Brown & Villareal, 2007) والتي أظهرت أن الطلبة منخفضي التحصيل يظهرون

استجابات غير مطابقة أكثر من الطلبة مرتفعي التحصيل في اختبار تكيفي في مادة الرياضيات وأرجعت الدراسة ذلك إلى انخفاض دافعية هؤلاء الطلبة وإلى لجوئهم للغش أكثر من الطلبة مرتفعي التحصيل الدراسي.

وتوصلت الدراسة إلى وجود اختلاف في عدد ونسبة الاستجابات غير المطابقة وفقا لمؤشري مربع البواقي على مستوى المفردة ومؤشري مربع البواقي البينية على مستوى أقسام الاختبار، حيث كانت نسبتها منخفضة وفقا لمؤشري مربع البواقي على مستوى المفردة مقارنة بتلك وفقا لمؤشر مربع البواقي البينية على مستوى أقسام الاختبار، ويمكن الاستنتاج من ذلك على أهمية استخدام مؤشرات على مستوى الأقسام لقدرتها في الكشف عن الاستجابات التي لا يمكن للمؤشرات على مستوى المفردة من الكشف عنها.

وبالنظر في قدرات الطلبة الذين يظهرون استجابات غير مطابقة وفقا لمؤشري مربع البواقي على مستوى المفردة، فقد كانت قدرات هؤلاء الطلبة منخفضة إلى متوسطة. كما كانت قدرات الطلبة الذين يظهرون استجابات غير مطابقة وفقا لمؤشري مربع البواقي البينية على مستوى أقسام الاختبار تتراوح بين المنخفضة والمتوسطة أيضا، وتشير قيم هذه القدرات إلى أن أغلبها مختلف بين الطريقتين، أي أن الأفراد الذين تم الكشف عن أن استجاباتهم غير مطابقة وفقا للطريقتين كانوا مختلفين، ويمكن القول أن عملية الكشف عن الاستجابات غير المطابقة تتطلب استخدام مؤشرات على مستوى المفردة وأخرى على مستوى أقسام الاختبار، كون الطريقتان تساعدان في الكشف عن أفراد مختلفين من ذوي الاستجابات غير المطابقة.

5. التوصيات والمقترحات:

في ضوء النتائج التي تم التوصل إليها، توصي الدراسة الحالية بالتالي:

1- استخدام مؤشر مربع البواقي البينية على مستوى أقسام الاختبار في الكشف عن الاستجابات غير المطابقة في اختبارات أخرى كاختبارات التنمية المعرفية واختبارات بيرلز لمقارنة خصائصه الإحصائية بما تم التوصل له في هذه الدراسة.
2- استخدام بيانات TIMSS2019 في العلوم وفي الصفين الرابع والثامن لدراسة الاستجابات غير المطابقة لطلبة سلطنة عمان ومقارنتها بما تم التوصل إليه مع بيانات الرياضيات في هذه الدراسة.
3- استخدام أقسام أخرى للاختبار للكشف عن الأفراد ذوي الاستجابات غير المطابقة مع مراعاة أن تكون هذه التقسيمات ذات معنى، وليس تقسيما عشوائيا وتوزيع الأسئلة على هذه الأقسام.
4- توجيه المعلمين والمعلمات والقائمين على تطبيق اختبارات TIMSS2019 بتدريب الطلبة على المهارات الرياضية الأساسية التي يحتاجونها لأداء هذه الاختبارات للتقليل من لجوئهم لسلوكيات تؤدي إلى استجابات غير مطابقة.
5- إجراء دراسات مقارنة في عدد الاستجابات غير المطابقة وفقا لمؤشر مربع البواقي المعيارية بين الدول المختلفة المشاركة في اختبارات TIMSS2019 من أجل التعرف على درجة التشابه والاختلاف وربطها بارتفاع وانخفاض مستويات الطلبة في الرياضيات والعلوم.
6- إجراء دراسات للوقوف على أسباب وجود الاستجابات غير المطابقة في اختبارات TIMSS2019 لدى طلبة سلطنة عمان.
7- إجراء المزيد من الدراسات عن تأثير عدد أقسام الاختبار، وعدد الأسئلة بكل قسم بالخصائص الإحصائية لمؤشري مربع البواقي البينية.

5. المراجع

1.5. المراجع العربية:

أبو عيش، بسينة رشاد بن علي. (2015). العوامل الشخصية والعادات الدراسية ذات العلاقة بتباين تحصيل الطلاب وطالبات الصف الثاني المتوسط في الرياضيات والعلوم في المملكة العربية السعودية في ضوء نتائج دراسة التوجهات الدولية للرياضيات والعلوم TIMSS2003. **المجلة العربية للعلوم الاجتماعية، المؤسسة العربية للاستشارات العلمية وتنمية الموارد البشرية، مصر، 16**(50)، 1-48.

بني عطا، زايد صالح إبراهيم. (2019). أنماط الاستجابة الشاذة في اختبار اوتيس - لينون المقنن للبيئة الأردنية وأثرها في دقة تقديرات قدرة الفرد ودالة المعلومات. **مجلة الدراسات التربوية والنفسية**، جامعة السلطان قابوس، 13(1)، 27-45.

الجراح، بندر نواف. (2020). الكشف عن أنماط الاستجابة في اختباري العلوم والرياضيات لبياناتTIMSS 2015 لدى عينة من طلبة الدولة السعودية ودولة سنغافورة باستخدام مؤشرات مطابقة الشخص. **مجلة العلوم التربوية**، جامعة الملك سعود، 23(2)، 299-320.

حمادنة، إياد محمد. (2005). تقصي فاعلية الإحصائي Lz^{new} في الكشف عن استجابة المفحوص غير المطابقة وفق نظرية استجابة الفقرة. **مجلة العلوم النفسية والتربوية**، جامعة آل البيت، 16(3)، 565-593.

زكي، مدحت. (2011). الاختبارات الدولية TIMSS. استرجع من:

https://taalmnashet.ahlamontada.net/t314-topic

شحادة، فوز حسن إبراهيم؛ والقراميطي، أبو الفتوح مختار. (2016). مستوى تحصيل طلبة المملكة العربية السعودية في الرياضيات والعلوم وفق نتائج الدراسات الدولية TIMSS مقارنة بالدول الأخرى من وجهة نظر المعلمين والمشرفين: الأسباب، الحلول والعلاج، أساليب التطوير. **مجلة التربية**، جامعة الأزهر، 169، 326- 370.

الشقصي، يعقوب؛ أبو شندي، يوسف؛ والمحرزي، راشد. (2019). فعالية مؤشرات مطابقة الفرد في نماذج الاستجابة للمفردة عند اختلاف قوة الارتباط الموضعي بين المفردات ونوع معالم النموذج. **مجلة الدراسات التربوية والنفسية**، جامعة السلطان قابوس،14(1)، 41-53.

الشمراني، صالح؛ الشمراني، سعيد؛ البرصان، إسماعيل؛ والدرواني، بكيل. (2016). إضاءات حول نتائج دول الخليج في التوجهات الدولية في العلوم والرياضيات TIMSS 2015. **هيئة تقويم التعليم**، جامعة الملك سعود.

عضيبات، تسنيم نعمان. (2019). **الممارسات التدريسية لمعلمي العلوم للصف الثامن في الأردن وسنغافورة في ضوء نتائج دراسة التوجهات الدولية للرياضيات والعلوم (TIMSS-2015)** [رسالة ماجستير غير منشورة]. جامعة اليرموك.

العلوي، سلطان بن ناصر بن سيف. (2017). **مدى تضمن موضوعات اختبار دراسة التوجهات الدولية في الرياضيات والعلوم في كتب العلوم للصفوف(5-8) بسلطنة عمان** [دراسة ماجستير غير منشورة]. جامعة السلطان قابوس.

عودة، أمل أحمد حمد (2018). **مقارنة فاعلية مؤشرات مطابقة الفرد حسب نماذج الاستجابة للفقرة ثنائية الاستجابة باستخدام مقياس الخليج للقدرات العقلية المتعددة.** [رسالة ماجستير غير منشورة]، جامعة السلطان قابوس.

النعيمي، عز الدين عبدالله عواد. (2006). **أثر انتهاك الاستقلال الموضعي على التقديرات المختلفة لنظرية استجابة الفقرة** [رسالة دكتوراه غير منشورة]. جامعة اليرموك.

2.5. المراجع الأجنبية:

Almehrizi, R. S. (2021). Coefficient beta as extension of KR-21 reliability for summed and scale scores. Applied Measurment in Education, 34, 139-149.
DOI: 10.1080/08957347.2021.1890740

Almehrizi, R. S. (2019). Residual-based person fit statistics over test sections. Journal of Educational and Psychological Studies, Sultan Qaboos University, 13(4),687-702.

Almehrizi, R. S. (2010). Comparison among new residual-based person fit indices and Wright's indices for dichotomous three-parameter IRT model with standardized tests. Journal of Educational and Psychological Studies, Sultan Qaboos University, 4(2), 14-26.

Almehrizi, R. S. (2004). Investigating a new modification of the residual-based person fit index and its relationship with other indices in dichotomous item response theory [Unpublished PhD Dissertation]. University of Iowa.

Belov, D. I., & Armstrong, R. D. (2010) Automatic detection of answer copying via Kullback-Leibler divergence and K- Index. Applied Psychological Measurement. 34(6), 379-392.

Biernbaum, M. (1986). Effect of dissimulation motivation and anxiety on response pattern appropriateness measures. Applied Psychological Measurement, 10, 167-174.

Brown, R., & Villareal, J. (2007). Correcting for person misfit in aggregated score reporting. International Journal of Testing, 7(1),1-25.

Bracey, Gerald & Rudner, Lawrence M. (1992). Person-fit statistics: high potential and many unanswered questions. Practical Assessment, Research & Evaluation, 3(7), 1-6.

Chen, J. (2004). Effect of test anxiety, time pressure, ability and gender on response aberrance. [Doctoral dissertation]. The Ohio State University.
https://etd.ohiolink.edu/apexprod/rws_etd/send_file/send?accession=osu1092840837&disposition=inline

D'Costa, A. (1993a, April). Extending the Sato caution index to define the within and beyond ability caution indexes. Paper presented at convention of National Council for Measurement in Education, Atlanta, GA.

D'Costa, A. (1993b, April). The validity of the W, B and Sato Caution indexes. Paper presented at the Seventh International Objective Measurment Conference, Atlanta, GA.

Felt,J. M., Castaneda, R., TiemensmaJ., & Depaoli, S. (2017). Using person fit statistics to detect outliers in survey research. Frontires Psychol., 8, 1-9.

Ferrando, Pere J.; Chico, Elise. (2001). Detecting dissimulation in personality test scores : A comparison between person-fit indices and detecting scales. Journal of Educational and psychological Measurment. Rovira i Virgili University, 61(6), 997-1012.

Glorfeld, L. W. (1995). An improvement on Horn's parallel Analysis methodology for selecting the correct number of factors to retain. Educational and Psychological Measurments, 55, 377-393.

Huang, T. (2011). Robustness of BW aberrance indices against test length. Knowledge Management & E-Learning: An International Journal, 3 (3), 310

Iasonas, C., Bill, B. & David, W. (2000). The consistency of examinee misfit across tests on the same subject and across subject: the case of the KS2 mathematics and science National Curriculum tests in England. Retrieved from: http//www.man.edu.uk

Drasgow, F., Levine, M. V., & Williams, E. A. (1985). Appropriateness measurement with polytomous item response models and standardized indices. British Journal of Mathematical and statistical psychology, 38, 67-86.

Karabatosos, G. (2003). Comparing the aberrant response detection performance of thirty-six person-fit statistics, Applied Measurement in Education, 16(4), 277-298.

Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). TIMSS 2019 Assessment Frameworks. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2019/frameworks/

Steinkamp, S. (2017). Identifying aberrant responding: Use of multiple measures.[Unpuplblished doctoral dissertation]. University of Minnesota. http://hdl.handle.net/11299/188885

Wright, B. D. (1977). Solving measurement problems with the Rasch model. Journal of Education Measurement, 114, 96-115.

TIMSS & PIRLS (n.d). International Study Centre. Boston College. Retrieved from: https://timssandpirls.bc.edu/index.html

جميع الحقوق محفوظة © 2024، الباحثة/ نوال علي اليحيائي، أ.د/ راشد سيف المحرزي، الدكتور/ إيهاب محمد عمارة،
المجلة الأكاديمية للأبحاث والنشر العلمي (CC BY NC)

Doi: https://doi.org/10.52132/Ajrsp/v5.59.7